



L'INFOMEDIATION DE L'INFORMATION EN LIGNE : LE CAS DES FILIALES FRANCAISES DE GOOGLE ET YAHOO

Robert Boure, Nikos Smyrnaio

► To cite this version:

Robert Boure, Nikos Smyrnaio. L'INFOMEDIATION DE L'INFORMATION EN LIGNE : LE CAS DES FILIALES FRANCAISES DE GOOGLE ET YAHOO. Document numérique et société, Sep 2006, France. pp.43-55. sic_00378766

HAL Id: sic_00378766

https://archivesic.ccsd.cnrs.fr/sic_00378766

Submitted on 26 Apr 2009

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

L'INFOMÉDIATION DE L'INFORMATION EN LIGNE Le cas des filiales françaises de Google et Yahoo

Si elle suppose une industrialisation avancée, l'information en ligne demande aussi la réunion de deux compétences différentes : une compétence éditoriale renvoyant largement aux médias et une compétence « technique », liée au support, qui renvoie à l'informatique. Cette dernière permet en partie de comprendre pourquoi des acteurs disposant initialement d'une forte expertise informatique ont pu progressivement s'engager sur le marché de l'information en ligne¹ avec un objectif clair : se positionner sur l'internet en tant qu'intermédiaires entre l'offre et la demande d'information en assurant une fonction s'apparentant à la distribution².

Google et Yahoo, déjà fortement présents dans d'autres domaines de l'industrie du Net (Smyrnaios, 2005), étendent désormais leurs activités à l'infomédiation de l'information en ligne, autrement dit à l'organisation et à la mise à disposition de documents numériques relatifs à l'actualité, triés et structurés par des algorithmes simulant une maîtrise de la sémantique. Cette extension, facilitée par la montée en puissance de l'internet à l'échelle mondiale ainsi que par le développement et le perfectionnement des technologies de recherche d'information, a pour conséquence principale le recul croissant des agents « humains » au profit de dispositifs socio-techniques prétendant assurer une infomédiation automatisée (Machill, Neuberger, Schweiger, Wirth, 2004). Selon Curien et Muet (2004, 40), cette dernière est d'autant plus institutionnalisée qu'elle est « *organisée par des sites commerciaux, des portails, ou des sites de médias cherchant à valoriser leur audience* ».

Trois caractéristiques majeures rapprochent Google et Yahoo : **1)** la fonction d'infomédiation est largement financée par la publicité ; **2)** ils se placent en position d'intermédiaires valorisant leur maîtrise des outils de recherche afin de proposer des pages web rassemblant des liens vers des informations situées sur d'autres sites. Ils organisent une offre hétérogène et éclatée en effectuant des rapprochements entre différentes sources sur la base de critères géographiques, sémantiques,

1 Par « information en ligne », nous entendons –sans préjuger de la pertinence d'autres définitions– les informations accessibles via des sites internet, mises à jour et renouvelées régulièrement (*news*) dont la production implique à un moment ou à un autre l'intervention de journalistes professionnels ou de rédacteurs spécialisés et qui sont destinées à un public large, indifférencié et anonyme. D'abord séparé de l'information scientifique et technique, ce type d'information est désormais partiellement relié à elle grâce à la numérisation et aux réseaux électroniques, mais aussi aux ponts qui s'établissent entre les mondes professionnels concernés.

2 Cela s'est manifesté de façon voisine pour les livres numérisés dans le cadre de la bibliothèque virtuelle Google Print (Jeanneney, 2005).

linguistiques ou chronologiques. Ce sont donc des acteurs concernés par l'assemblage et l'édition de contenus. Selon Bustamante (2005, 807), cette situation déjà observée dans la télévision payante par satellite ou par câble, « *s'étend maintenant dans tous les types de services culturels et d'information. En tous les cas, cette fonction apporte une valeur ajoutée, mais en même temps, absorbe une partie des profits, ce qui génère des tensions entre créateurs, producteurs et fournisseurs de contenu* » ; **3)** contrairement aux fournisseurs d'accès, leur activité ne nécessite pas, du moins au départ, d'importants et coûteux investissements « matériels » (par exemple, la mise en place et l'entretien de réseaux de télécommunications). De sorte qu'ils peuvent offrir de nombreux services par le biais de leurs sites portails sans pour autant s'implanter fortement à l'étranger. Et quand ils s'y installent à travers des filiales dédiées, il s'agit, en tout cas dans un premier temps, d'une implantation essentiellement commerciale.

Mais par-delà ces ressemblances, ces acteurs ont des approches différentes de l'information en ligne, phénomène perceptible tant au niveau de la conception qu'à ceux de l'organisation et du fonctionnement des sites. Ces différences s'expliquent par leurs conceptions de l'internet, leur histoire, leurs avantages concurrentiels, mais aussi leurs positionnements stratégiques (I). Elles permettent de mieux comprendre les dispositifs d'offre de leurs sites français d'information en ligne dédiés à l'actualité dont nous dresserons une typologie avant d'examiner les hiérarchisations de l'information qu'ils reproduisent et/ou construisent (II).

I) DES POSITIONNEMENTS STRATÉGIQUES DIFFÉRENTS

1) Sous les moteurs de recherche... les stratégies

Les stratégies de Google et Yahoo prennent appui sur un socle de services de base liés à l'utilisation du réseau dont une des composantes majeures est la recherche d'informations.

Dès 1996, Sergey Brin et Larry Page, fondateurs de Google, ont conçu le principe d'un moteur de recherche ne fonctionnant pas exclusivement sur la base des mots-clés présents sur une page web et prenant en compte le nombre et la qualité des liens externes qui pointent vers la même page (Hindman, Tsioutsoulis, Johnson, 2003). Chaque lien hypertexte reliant deux sites est interprété par l'algorithme *PageRank*, propriété de Google, comme un « vote » de l'un au profit de l'autre. Ainsi, la pertinence d'une page web concernant une requête est évaluée non seulement par rapport à l'existence des mots-clés recherchés, mais également par rapport au nombre de « votes ». Google reproduit d'une certaine manière une hiérarchie déjà établie sur l'internet par les gestionnaires des sites qui renvoient vers d'autres pages web qu'ils considèrent intéressantes.

Au départ, Yahoo fonctionnait sur un principe différent, l'annuaire. Les documentalistes de la société établissaient manuellement des catégories de sites web afin d'organiser les documents

disponibles sur la base de critères thématiques. Parallèlement, jusqu'au début des années 2000, la recherche d'information automatisée était sous-traitée auprès de Google. Puis, en raison du succès économique des moteurs de recherche, Yahoo a opéré un virage stratégique en rachetant Inktomi (2002), fournisseur des technologies de recherche d'information, et en suspendant l'accord le liant à Google.

De fait, l'amélioration de l'efficacité de leurs outils de recherche constitue désormais une préoccupation stratégique majeure des acteurs engagés dans le marché de l'infomédiation. Ils estiment que cette amélioration passe, entre autres, par l'intégration dans le dispositif technique des motivations et des pratiques des usagers. Il s'agit alors d'appliquer les algorithmes de pertinence non plus à l'ensemble des pages web recensées par leurs robots respectifs, mais à un « corpus » de contenus constitué par des sites-sources présélectionnés en fonction d'un certain nombre de critères.

2) Le positionnement « média » de Yahoo

En matière d'information de presse, Yahoo occupe un positionnement « média ». En effet, il s'efforce de se raccrocher au système médiatique du pays dans lequel il s'installe en établissant des relations avec les éditeurs traditionnels. Par ailleurs, à sa direction américaine, mais aussi à la tête de ses filiales nationales, Yahoo emploie des salariés dont le parcours professionnel s'est essentiellement déroulé dans le secteur des industries de la culture et de la communication³. Vis-à-vis des usagers, sa stratégie est de mettre à disposition une production d'informations en partie originales, constituée d'articles, d'images vidéos, d'images fixes et de sons. Pour cela, il fait appel à des agences spécialisées auxquelles il commande des contenus d'information originaux (ex : formats originaux combinant texte et vidéo de l'agence Digipresse), à des journalistes indépendants (Kevin Sites et sa rubrique Hotzone dans la version américaine) et à des intervenants ponctuels (éditoriaux de Bernard Cassen, président d'honneur d'ATTAC lors de la campagne référendaire sur la Constitution européenne dans la version française). Dans la même logique, Yahoo France commande des sondages d'opinion sur des questions d'actualité à des instituts spécialisés, en partenariat avec d'autres médias et notamment *Libération*. Cette politique implique une implantation « non symbolique » dans les pays où elle est appliquée, et notamment en France où la filiale compte 70 salariés, dont une partie significative est en charge des relations avec les éditeurs de contenus et de services. Elle suppose aussi une stratégie prenant en compte la spécificité de chaque système médiatique national.

Le positionnement « média » nécessite également un travail manuel conséquent de la part des

³ Ainsi, Terry Semel, PDG de Yahoo, a été pendant longtemps dirigeant de majors du cinéma comme Warner Bros ou Walt Disney. John Marcom, responsable des activités à l'international, est un ancien journaliste du *Wall Street Journal* ayant d'abord travaillé pour Time-Warner. Maxime Zeller, responsable éditorial du portail yahoo.fr, vient du groupe de presse *L'Express-L'Expansion*.

gestionnaires du système technique d'information. Ils doivent paramétrer ce dernier afin d'effectuer un *cross-linking* efficace, c'est-à-dire une mise en relation des différents contenus sur une base thématique en utilisant des critères fondés sur les meta-données (mots-clés, rubriques) qui accompagnent les articles, images, sons et vidéos. Le même processus est utilisé pour le *full coverage*, autrement dit les dossiers thématiques régulièrement créés sur des questions d'actualité⁴. Les responsables du service Actualité réalisent aussi un travail d'écriture vis-à-vis de l'accroche de la rubrique sur la page d'accueil du portail. Cela permet de donner un ton particulier au service et de distiller de « l'impertinence » et de « l'humour » (cf. entretiens).

On notera enfin que le modèle publicitaire de Yahoo est proche d'un positionnement « média ». En effet, si la société est engagée dans le marché des liens sponsorisés, notamment depuis le rachat de Overture, l'essentiel de ses recettes provient de la vente aux annonceurs des espaces publicitaires du portail. D'où la mise en place d'offres spécifiques telle *Ad Premium CSP+* qui propose d'une part, un espace exclusivement dédié à l'annonceur sur la page d'accueil toute la journée du lundi, d'autre part, des espaces ciblés dans les différentes rubriques éditoriales le reste de la semaine (<http://fr.adinfo.yahoo.com/adpremium.html>). Le lundi matin est considéré par la régie publicitaire comme l'équivalent du *prime time* de la télévision en raison de son audience massive sur les lieux de travail, surtout chez les cadres à fort pouvoir d'achat. Elle n'hésite pas à comparer, en termes d'efficacité, son « offre premium » aux matinales des radios d'information ou aux quatrièmes de couverture des magazines grand public et des quotidiens d'information (cf. entretiens).

3) Le positionnement « moteur de recherche » de Google

Google exploite sa position dominante dans le domaine de la recherche d'informations en prolongeant ses activités, et par exemple en créant dans chaque pays des services annexes tel Google Actualités dans le cas de la France⁵. Le fonctionnement « matriciel » de la firme lui permet de concentrer ses divisions stratégiques aux États-Unis et de laisser aux équipes nationales (de taille très réduite) la gestion de la fonction de commercialisation des offres publicitaires. Ainsi, Google Actualités a été lancé directement depuis les États-Unis et est toujours géré depuis ce pays, sans que la filiale française puisse exercer un réel contrôle sur son fonctionnement. En 2005, l'équipe française compte moins de 30 salariés, pour l'essentiel des commerciaux. Les conséquences de cette absence d'intégration au système médiatique français sont multiples tant au niveau éditorial (quantité et diversité des informations, choix et diversité des sources, sélection des informations, absence de rédacteurs professionnels de l'information journalistique...) qu'à celui des relations avec les producteurs nationaux d'information (par

⁴ Les dossiers thématiques qui ont la plus grande audience traitent de questions déjà à la Une des médias : guerre en Irak en 2003, élections présidentielles aux États-Unis en 2004, référendum français sur la Constitution européenne, crise du CPE (cf. entretiens).
⁵ En 2004, 74% des recherches en France ont été réalisées grâce à Google (Panorama Médiamétrie-eStat/@position, *Bilan 2004 des indicateurs clés de l'Internet*, http://www.mediametrie.com/resultats.php?resultat_id=95&rubrique=net)

exemple, en matière de droits d’auteur)⁶.

Le principe qui anime Google News, dont Google Actualités est la déclinaison française, semble proche de celui d’un agrégateur d’informations sur l’actualité comme Yahoo Actualité. Il s’agit de constituer une page web sous la forme de Une d’un journal électronique découpée en rubriques thématiques (« International », « France », « Economie », « Science », « Sport »,...). Sous chaque rubrique thématique se trouve un ensemble de titres et de « chapeaux » d’articles, mais aussi de photos, sous forme de liens hypertextes conduisant vers le texte intégral. Le service, disponible en 22 versions nationales, comprend plusieurs milliers de sources.

La grande différence avec les pages d’information des portails généralistes, c’est que les articles ne se trouvent pas sur Google, mais sur les sites d’origine. Autrement dit, alors que les agrégateurs classiques sinon rémunèrent, du moins établissent des partenariats avec les producteurs de contenus afin de disposer du droit de les publier, Google ne fait que « pointer » vers des contenus externes pour lesquels il ne dispose d’aucun droit de publication. En exploitant sa technologie de recherche, il peut sonder en continu un grand nombre de sources présélectionnées et proposer des pages d’information à jour sans s’entendre avec les producteurs d’informations, et *a fortiori* les rémunérer. Ce système est censé être profitable aux deux parties (cf. entretiens) : Google utilise les contenus afin d’augmenter l’audience de son portail et multiplier ses recettes publicitaires grâce aux liens sponsorisés, tandis que les éditeurs bénéficient d’une audience accrue.

Les créateurs de Google News ont toujours valorisé l’aspect automatique et algorithmique du traitement, de la sélection et de la hiérarchisation de l’information. De leur point de vue, cette caractéristique confère au service un caractère « objectif » et « neutre » car elle permet d’éviter les biais médiatiques traditionnels (ligne éditoriale engagée, positionnement politique explicite..). De plus, Google News serait un service apportant une certaine exhaustivité quant à la couverture de l’actualité, ce qu’aucun média traditionnel ne peut faire (Ulken, 2005).

Enfin, le financement de Google provient essentiellement de la vente des liens sponsorisés, placés sur son site à côté des résultats de recherche. Google a également mis en place la régie publicitaire AdSense. Il s’agit d’un dispositif complètement automatisé qui attribue des publicités aux sites adhérents après avoir analysé leur contenu. En principe, cette méthode permet de ventiler les espaces achetés par des annonceurs au sein des sites comportant des contenus en rapport avec le message publicitaire, ce qui augmente l’efficacité des annonces. Le dispositif fonctionne sans intervention humaine, ce qui entraîne des économies substantielles par rapport au

⁶ En 2003, avant qu’un accord ne soit trouvé, Google a été assigné en justice par certains membres du GESTE (Groupement des éditeurs de services en ligne) pour utilisation abusive de leurs contenus. De même en mars 2005, l’Agence France Presse l’a attaqué en justice en l’accusant d’avoir violé le *copyright* sur ses dépêches et ses photos. Avant le jugement, Google a retiré de son site le contenu AFP.

fonctionnement d'une régie publicitaire « classique » (Rothschild, 2005).

II) TYPOLOGIE ET HIERARCHIE DES SOURCES ET DES INFORMATIONS

La sélection, le classement et la hiérarchisation des sources et des informations sont des activités bien connues de la presse. Comment les infomédiaires « font-ils avec » elles ?

L'infomédiation assurée par les services d'actualité de Google et Yahoo consiste à donner aux usagers accès à des informations produites par des tiers. La question qui se pose est de savoir quelles sont ces sources et ces informations et comment fonctionne le système de sélection et de hiérarchisation. Afin d'y répondre, nous avons été contraints de recourir, pour chaque acteur, à un système de collecte de données différent, ce qui rend les comparaisons point à point impossibles, mais autorise néanmoins la mise en évidence de différences à partir d'observables.

Nous avons exploré Yahoo Actualités au moyen du robot d'extraction Zoom. Le robot a été réglé à un niveau de profondeur de « un clic » à partir de chaque rubrique (« A la Une », « Monde », « France », « Eco », « Multimédia », « People », « Insolite », « Sport »...). Ces opérations ont été réalisées le 3 avril 2006. Zoom n'étant pas utilisable pour explorer Google, nous avons eu recours aux données produites par Google News Report et disponibles, pour la version française, sur le site www.privateradio.org. Il s'agit de données quantitatives obtenues grâce à un robot d'extraction visitant à intervalles réguliers la Une et extrayant des informations relatives notamment aux sources et aux articles. Pour chaque acteur, cette recherche a été complétée par des entretiens non directifs avec des responsables commerciaux ou stratégiques.

1) Yahoo Actualités

Yahoo Actualités fonctionne sur la base d'une plateforme modulable qui réceptionne toutes les unités d'information, puis organise et édite les pages. Si l'on met à l'écart les archives, l'information en ligne se divise en trois catégories :

- **les contenus venant de partenaires.** Ce sont des flux d'informations sous format NewsML, dérivé du méta-langage Extensible Markup Language (XML)⁷. Les partenaires de Yahoo sont soit des fournisseurs classiques comme les trois grandes agences de presse AFP, AP et Reuters, soit des éditeurs de sites qui s'adonnent à un échange « trafic contre contenu ». Les premiers sont rémunérés afin de mettre à disposition du client leurs fils de dépêches, alors que les seconds

⁷ XML est un standard du World Wide Web Consortium permettant de créer des langages de balisage. Ces derniers sont des langages spécialisés dans l'enrichissement de l'information textuelle. Ils opèrent grâce aux balises, unités sémantiques délimitant un ensemble à l'intérieur d'un fichier texte. L'inclusion de balises permet de transférer la structure du document et son contenu. XML autorise donc le partage de textes et d'informations structurées en séparant le contenu (les données) du contenant (leur présentation).

fournissent du contenu en échange des liens renvoyant vers leurs sites respectifs, profitant ainsi de l'audience de Yahoo.

Les contenus ont des formes variées (textes, photos, infographies, vidéos, sons, liens hypertexte). Tous contiennent des méta-données définies par le fournisseur, informations censées permettre à Yahoo d'éditer ses rubriques d'actualités « de façon efficace » (cf. entretien). Outre les titres et sous-titres des articles, ces méta-données comportent la date et l'heure de la publication, des mots-clés précisant les thématiques, la catégorie dans laquelle le fournisseur a rangé l'information parmi ses différentes rubriques. Dans le cas des agences de presse, le flux XML comporte également une hiérarchie des informations définie par le fournisseur.

- **les sources externes.** Il s'agit d'une base de données alimentée manuellement par les documentalistes de Yahoo⁸. Ces sources peuvent être rangées essentiellement en quatre catégories : a) les sites-médias français et étrangers, autrement dit des publications en ligne éditées par des médias (stations de radio, chaînes de télévision, presse écrite) ; b) les publications –souvent spécialisées– spécifiques à l'internet, c'est-à-dire provenant de sites dont le contenu est mis à jour régulièrement par des journalistes ou des rédacteurs ; c) les blogs ; d) les sites institutionnels.

Cette base de données, fondée en grande partie sur l'annuaire de Yahoo (un des plus importants de l'internet), peut aussi être ponctuellement enrichie si nécessaire. Cependant, l'activité historique de recensement manuel des sites pour l'annuaire est peu à peu reléguée au second rang, notamment au niveau de « l'affectation des ressources humaines » (cf. entretiens).

L'observation réalisée avec Zoom fournit des informations utiles sur ces sources externes. 378 sources ont été recensées pour une seule journée (deux autres observations réalisées à deux dates différentes donnent des résultats très voisins) : 28,8% de sites-médias (68 sont français, 21 africains –Maghreb et Afrique francophone surtout–, 10 européens, 5 nord-américains et 5 sud-américains), 13,7% de publications spécifiques au réseau et 7,4% de blogs. 50% des sources externes (soit 189) est donc de nature institutionnelle (associations, collectivités publiques, entreprises). Leur présence permet aux usagers de compléter l'information fournie par les agences de presse et les médias et d'approfondir les questions d'actualité en renvoyant vers les sites des organisations publiques et privées concernées.

- **les autres services de Yahoo.** Le large éventail de services dont dispose Yahoo peut à tout moment enrichir les pages d'actualité. Un exemple significatif est la mise à disposition de Flickr, service de partage et de classement de photos. Flickr est un service initialement créé par une

⁸ Les liens vers les sources externes sont de deux types : liens « classiques » menant vers la page d'accueil des sites et liens « profonds » donnant accès à des contenus spécifiques dans les pages web.

société canadienne (rachetée par Yahoo en 2005) afin de permettre aux usagers de publier leurs photos en ligne tout en paramétrant les dispositions d'accès. Un système de classement sur la base de mots-clés donne la possibilité de créer des collections thématiques en regroupant des photos prises par des personnes différentes. Tout au long du mouvement anti-CPE, des centaines d'internautes ont publié ainsi leurs photos des événements. Le soir de la dernière manifestation nationale (28 mars 2006), Flickr proposait plus de 3 000 clichés sur le sujet, dont plusieurs centaines relatifs au 28 mars. Un onglet renvoyant vers ces photos était mis en place sur chaque page de Yahoo Actualités consacrée à ce sujet.

Ces trois types d'informations en ligne sont traités quotidiennement par la plateforme modulable de gestion de contenus. Le système analyse les méta-données attachées aux unités d'information et édite de façon automatique les pages web. Il est à noter que l'édition des pages d'information reprend toujours la même architecture : les contenus achetés aux partenaires, essentiellement les agences de presse, occupent une position centrale dans les pages d'actualité courante et les dossiers thématiques.

La plateforme doit gérer et mettre en relation plusieurs flux de contenus représentant des volumes importants. A l'inverse des portails des FAI qui reprennent en l'état la hiérarchie de l'information établie par les agences, Yahoo se réserve la possibilité de répartir en rubriques et de hiérarchiser les sujets en fonction de ce qu'il faut bien appeler une « ligne éditoriale ». Dès lors il peut y avoir opposition ou adéquation entre les grilles de hiérarchisation et de catégorisation des fournisseurs et de Yahoo. Le travail des responsables de Yahoo Actualités consiste à construire une grille comprenant des catégories, des mots-clés et des opérateurs booléens qui suivront la dépêche et lui feront intégrer automatiquement la ou les rubriques choisies, et à l'intérieur de celles-ci, une certaine place. Ces critères s'efforcent d'anticiper les points d'accès possibles des usagers à une thématique. C'est la raison pour laquelle la plateforme de Yahoo est modulable : elle combine, selon les besoins, traitement automatique de l'information et intervention manuelle des responsables. Mais depuis 2002, la tendance est à la diminution du travail humain au profit du fonctionnement automatique. Ainsi, le nombre de personnes éditant les pages de Yahoo Actualités est passé de trois à une. Dès lors l'activité du préposé consiste essentiellement en la maintenance de la plateforme et en la rédaction d'une accroche quotidienne pour la page d'accueil.

2) La version française de Google News

La version française de Google News fonctionne de manière totalement automatisée. Initialement, un documentaliste basé aux États-Unis était chargé de constituer une base de données comportant des sources d'information en français et pouvant être modifiée au fur et à mesure. Désormais, le moteur de Google pointe les sites-sources de façon régulière, environ

toutes les quinze minutes, et exploite le contenu de deux manières : il crée la Une de Google Actualités en distribuant les liens vers les articles des différentes rubriques et il fournit une liste de réponses relatives à des requêtes effectuées depuis l'interface de recherche du moteur.

La liste des dix premiers sites-sources (sur un total de 810) ayant produit 19635 articles sur sa Une (période du 1er janvier au 3 avril 2006) fait apparaître une forte présence des grands médias français, et spécialement de la presse nationale : *Le Nouvel Observateur* (5430 articles), *Libération* (2292), *L'Express* (2110), *Le Figaro* (1783), Reuters (1740), *La Tribune* (1537), Boursier.com (1500), *Le Monde* (1224), Xinhua, agence de presse chinoise (1184) et *TF1* (835). Cependant, alors que 59 sources sont mises en avant de façon systématique (plus de 100 articles sur les trois mois, soit plus d'un par jour), 650 ont moins de 20 articles référencés. Parmi les 59 sources dominantes, 30, 5% sont des sites-médias français, 22% des sites-médias étrangers, 5% des agences de presse... et 42% des sites exogènes à la sphère des médias⁹.

La hiérarchisation de l'information s'effectue sur la base d'un système complexe construisant des regroupements d'éléments, en l'occurrence des articles et des photos, sur un même thème ou sur une même rubrique. Cependant, les données que nous avons recoupées indiquent qu'il s'agit d'une combinaison entre l'algorithme *PageRank*, fondé sur le nombre de liens pointant vers la même page web et d'autres critères comme le contenu du texte, des titres et sous-titres, la date et l'heure de la publication... Il semble y avoir une hiérarchisation préalable –et opaque– des sources qui privilégie certaines d'entre elles pour la construction de la Une. Tout se passe comme si Google voulait prendre en compte le « capital journalistique » des sites-sources, c'est-à-dire celui qui correspond à « *une reconnaissance qui s'obtient auprès de ses pairs, mais aussi auprès du public et de pouvoirs externes* » (Duval (2004, 107). Est-ce possible, car il ne cultive pas ses relations avec le monde des médias et le capital journalistique est une alchimie sociale difficilement traduisible en algorithme ?

Conscient du problème, il s'efforce d'intégrer à son algorithme de hiérarchisation des éléments visant davantage à simuler le capital journalistique qu'à le prendre réellement en compte. Aucune information précise ne filtre sur la méthode car la société est sensible sur tout ce qui concerne la « boîte noire » de sa technologie. Cependant, un brevet a été déposé aux États-Unis et un article de *New Scientist* permet d'y voir plus clair sur ses intentions (Fox, 2005). On y apprend qu'il cherche à constituer une base de données permettant de comparer la crédibilité des différentes sources d'information et de modifier en conséquence leur classement. Les critères révélés sont le nombre d'articles produits par chaque source et leur taille moyenne, l'ancienneté des sites, leur

9 Les sites exogènes sont très présents dans les rubriques spécialisées du type « Sciences/Technologie », « Santé », « Sport », alors que les sites-médias et ceux des agences de presse dominent la Une ainsi que les rubriques généralistes « France » et « Monde ». Cela est lié au fait que les premiers assurent la diffusion gratuite et en flux continu de brèves renvoyant à l'actualité des rubriques spécialisées : ainsi, les sites Sports.fr (541 articles) et Sport24 (278 articles), peu connus, arrivent très loin devant celui de *L'Equipe* (74 articles) qui constitue la référence en matière sportive dans le système médiatique français.

nombre d'employés, le volume de leur trafic et le nombre d'internautes qui y accèdent à partir d'autres pays. Sont-ils pertinents ? Quel sera leur poids relatif dans l'algorithme ?

Un début de réponse est donné par un élément essentiel dans la hiérarchisation de l'information sur la Une : la « fraîcheur » des articles. En effet, la particularité de l'actualité résidant dans son caractère éphémère, le jour et l'heure de la publication d'un article revêtent une importance capitale. En somme, plus un article est récent et plus la source dispose d'un capital journalistique important, plus il a de chances d'être mis en avant sur Google News. Ce mode de fonctionnement aboutit à la valorisation quasi-systématique des sources qui publient des informations à un rythme soutenu, privilégiant ainsi les sites dont les rédactions sont organisées dans l'objectif d'un renouvellement perpétuel du contenu. On relèvera que la première source mise en avant par la version française est de loin le site du *Nouvel Observateur* (cf. *supra*). C'est moins un effet de la crédibilité du média que de l'organisation interne de son site *nouvelobs.com* conçue en totale cohérence avec le fonctionnement des infomédiaires comme Google News. En effet, la rédaction internet de l'hebdomadaire est constituée de jeunes journalistes dont le travail consiste pour l'essentiel en la rédaction à flux tendus d'articles à partir des dépêches d'agences afin de les publier immédiatement sur le site. En lui fournissant en continu de quoi alimenter son service d'actualité, *Le Nouvel Observateur* exploite efficacement, en termes d'audience, ce nouveau canal de distribution qu'est Google¹⁰. Progressivement, les sites-médias mentionnés *supra* adoptent peu ou prou cette production à flux tendu d'articles, dont un grand nombre ne sont en réalité que des synthèses de dépêches d'agences.

On complétera cette analyse en mentionnant le système de personnalisation proposé par Google. Les usagers ont la possibilité d'avoir « leur » Une. Outre la réorganisation des rubriques de la page d'accueil (par exemple, en intégrant des rubriques en langues étrangères), ils peuvent personnaliser la recherche relative à l'actualité. Il leur suffit de s'inscrire auprès de Google avec un nom d'utilisateur et un mot de passe et d'autoriser l'infomédiaire à cumuler des informations relatives à ses requêtes et ses consultations de contenus afin de construire un profil fondé sur ses centres d'intérêt. Ces paramètres sont ensuite intégrés dans Google Actualités qui modifie sa hiérarchie au profit de la Une personnalisée. Ainsi, une source très régulièrement consultée par un usager sera davantage mise en avant dans la Une personnalisée que dans la Une standard.

Cette recherche, centrée sur les deux infomédiaires de l'actualité les plus importants en termes d'audience, a mis en évidence des différences sensibles au niveau des stratégies et des manières de faire, même si quelques points de convergence ont pu être relevés¹¹. Alors que Yahoo cherche,

¹⁰ Il est d'ailleurs un des médias à ne pas s'être associé aux actions entreprises par le GESTE (2003) contre Google et à avoir très tôt établi avec lui des relations commerciales lui en achetant des liens sponsorisés.

¹¹ Nos conclusions sur la version française Google concordent avec celles d'autres travaux relatifs à la version allemande (Schroeder, Kralemann, 2005)

tout en faisant valoir ses spécificités, à intégrer le système médiatique national, notamment au travers de relations multiples avec ses acteurs traditionnels, Google se superpose à lui en tentant de devenir un acteur incontournable. Pour les deux se pose, mais en des termes différents, les questions de leur complémentarité et de leur concurrence avec les acteurs médiatiques traditionnels.

Cependant, en raison de la montée en régime de l'infomédiation, émergent d'autres acteurs tels Wikio ou Newsisfree, totalement spécialisés dans l'actualité en ligne, arrivent sur le marché avec d'autres stratégies et d'autres méthodes. Il y a là matière à de nouvelles recherches...

Curien (N.), Muet (P.-A.), 2004, *La société de l'information*, Rapport au Conseil d'analyse économique, Paris : La Documentation française.

Duval (J.), 2004, *Critique de la raison journalistique, les transformations de la presse économique en France*, Paris : Le Seuil.

Fox (B.), 2005, « Google searches for quality not quantity », *New Scientist Magazine*, n°2497, 30 avril.

Hindman (M.), Tsioutsoulis (K.), Johnson (J.A.), 2003, « Googlearchy : how a few heavily-linked sites dominate politics on the web », *Annual Meeting of the Midwest Political Science Association*, Chicago, 4 avril. <http://www.cs.princeton.edu/~kt/mpsa03.pdf>.

Jeanneney (J.-N.), 2005, *Quand Google défie l'Europe. Plaidoyer pour un sursaut*, Paris : Mille et une nuits.

Machill (M.), Neuberger (C.), Schweiger (W.), Wirth (W.), 2004, « Navigating the internet. A study of german-language search engines », *European Journal Of Communication*, vol. 19 (3), 321-347.

Rothschild (M.), 2005, « Network externalities, search engines, and advertising or what is Google worth », colloque *The economics of the software and internet industries*, Toulouse, 21 et 22 janvier.

Schroeder (R.), Kraleman (M.), 2005, « Journalism ex machina – Google News Germany and its news selection processes », *Journalism Studies*, vol. 6, n°2, 245-247.

Smyrniaios (N.), 2005, *L'industrie de l'éphémère. Emergence et consolidation de modèles diversifiés de production et de diffusion de l'information en ligne*, Thèse en Sciences de l'information et de la communication, Université Grenoble 3.

Ulken (E.), 2005, « A question of balance : are Google News search results politically biased ? », *USC Annenberg School for Communication Review*, vol. 2(3), may, 78-99.